DOCUMENT RESUME

ED 364 417 SE 053 896

AUTHOR Millar, Robin; And Others
TITLE Children's Performance of Investigative Tasks in Science: A Framework for Considering Progression.
PUB DATE Sep 93
NOTE 32p.; Paper presented at the Annual Meeting of the British Educational Research Association (Liverpool, England, United Kingdom, September 10-13, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Concept Formation; Educational Research; Elementary School Science; Elementary School Students; Elementary Secondary Education; Foreign Countries; *Inquiry; Process Education; Science Education; Scientific Concepts; Secondary School Science; Secondary School Students; *Student Evaluation
IDENTIFIERS *PACKS Project (United Kingdom); *Performance Based Evaluation; Progress Reporting; Science Process Skills; United Kingdom

ABSTRACT

This paper begins by exploring some general issues associated with the idea of progression in education in general and in science education in particular. These are then related to the particular domain of interest of the Procedural and Conceptual Knowledge in Science (PACKS) Project, which is concerned with the responses of children, aged 9 to 14, to science investigation tasks and the understanding they bring to bear in responding. A model linking performance and understanding is proposed. Data from PACKs Project is used to illustrate one aspect of this model, showing how decisions about investigation frame provide one indicator of progression in performance. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 364 417

Children's Performance of Investigative Tasks in Science: a framework for considering progression

Robin Millar[1]
Richard Gott[2]
Fred Lubben[1]
Sandra Duggan[2]

1 :
Department of Educational Studies
University of York
Heslington
YORK YO1 5DD
UK

2 :
School of Education
University of Durham
Leazes Road
DURHAM DH1 1TA
UK

Abstract:

This paper begins by exploring some general issues associated with the idea of progression in education in general and in science education in particular. These are then related to the particular domain of interest of the *Procedural and Conceptual Knowledge in Science (PACKS) Project*, which is concerned with the responses of children, aged 9 to 14, to science investigation tasks and the understanding they bring to bear in responding. A model linking performance and understanding is proposed. Data from the PACKS Project is used to illustrate one aspect of this model, showing how decisions about *investigation frame* provide one indicator of progression in performance.

Paper presented at the annual meeting of the British Educational Research Association, University of Liverpool, 10-13 September 1993.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

P. H. Millar

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Introduction

The *Procedural and Conceptual Knowledge in Science (PACKS)* project is concerned with the responses of children, aged 9 to 14, to given science investigation tasks. Investigations, as a component of science teaching and learning, have grown in prominence in recent years in the UK and are now established as a major element of school science within the National Curriculum model (DES 1989, 1991). Whilst there is widespread agreement amongst science teachers and educators that developing the ability to carry out a scientific investigation is an important part of learning science, there is no established tradition of teaching or assessing investigation performance. As a result, National Curriculum Attainment Target Sc1 Scientific Investigation has caused considerable difficulty for many teachers and has stimulated more discussion and criticism than most other aspects of the Science National Curriculum.

The focus of interest in the PACKS project is on the understandings which underpin children's performances in response to investigation tasks. We are interested in the interplay between children's understanding of the science facts, ideas and concepts which are relevant to a given investigation (sometimes referred to as *conceptual knowledge*), their understanding of the procedures of scientific investigating (*procedural knowledge*), and the details of their actual task performance. The aim of the project is to develop and validate a model relating aspects of understanding to elements of performance.

Such a model would allow us to begin to characterise more precisely the nature and extent of the differences between "novice" and "expert" performance in investigating in science. That is, it would provide the ground for a model of progression in performance. This paper begins by exploring some general issues associated with the idea of progression. It then relates these to the particular domain of interest to the PACKS project, and proposes a model linking understanding and performance. Data from the project are then used to illustrate differences in performance in relation to this model. The paper ends with some general comments about the nature of progression in performing science investigations.

The practical importance of a model of progression

Some notion of progression, either implicit or explicit, underpins all teaching, and all syllabus and course construction and assessment. Consider the "history" of an individual child's experience of a school subject, such as science. She or he will experience a sequence of science activities, either embedded in topics or explicitly separated out as "science lessons", planned by a series of teachers. The teacher provides some experience or activity for the child in each lesson or "time slot"; the child experiences activities and events in a time-sequence. The choice for the teacher (or for whoever plans the programme of instruction) is not between *having* and *not having* a model of progression; it is between having an *implicit* one or one which is, to some extent at least, *explicit*.

So, in thinking and talking about progression, we are asking whether there are better and worse ways of organising the time-series of experiences presented to children. Teachers have a great deal of intuitive "knowledge" about progression, largely based in experience and tradition (including their own experience of being pupils, as well as their teacher training and socialisation as teachers). They have a "feel" about what is appropriate at certain ages or stages, about what comes before what, about what should be delayed until later, and so on. Some of these ideas are based on views about children as learners, others on views of the subject matter. Apart from its intrinsic interest, research on progression is important because it may lead to other bases, more securely grounded in empirical data or theory, which can provide better guidelines for teachers (and others, such as textbook authors and curriculum planners) in making decisions about the actual sequence of events and experiences presented to the child.

Progression of tasks or progression of learners?

Progression may be thought of in terms of the increasing demand of the tasks learners are given, or in terms of the increasing sophistication of responses to similar tasks. The former focuses on the nature of learning tasks. Syllabuses and programmes of study frequently embody such notions of increasing *task demand*. The latter focuses instead on the varying qualities of learners'

responses to essentially similar tasks, seeing progression in terms of *task performance*. This sense of progression is particularly important in subjects, or domains, where tasks are relatively open-ended, permitting a wider range of possible responses. The distinction between progression in *task demand* and in *task performance* is used extensively by Qualter *et al* (1990) to structure their discussion of Sc1. These two senses of progression are, however, related through the single notion of progression in a *learner's capabilities*. This will be reflected in:

- ability to perform the same (or similar) tasks in a more "expert" manner; and
- ability to perform *more demanding* tasks.

In the PACKS project we are interested to explore the nature of the understandings which lead to, or facilitate, progression in learners' capabilities in scientific investigating.

Models of progression in science

In science, it is useful to think about progression in understanding in distinct content domains, such as electric circuits, or forces and motion ("knowing what"); or progression in the ability to carry out specific tasks, such as setting up and using a microscope to look at cells ("knowing how")[1]. In the former, the aim is greater "understanding", in the latter we sometimes talk of greater "skill". There is, however, less difference between "understanding" and "performance" domains than is sometimes claimed (or implied). First, it is clear that many "skill"-type performances depend upon having some knowledge and understanding. This is particularly so in the case of (so-called) "skills" such as "designing an experiment", but is also true in the example given above, setting up a microscope. Second, and perhaps more fundamentally, the only way we have of assessing a learner's "understanding" is through some observable performance, such as their written or spoken answer to a diagnostic question, or their ability to engage appropriately in written or spoken dialogue about some

[1] The terms "knowing what" and "knowing how" come from Ryle (1949).

phenomenon within the domain. That is, we have no direct access to "understanding"; we infer it from specific features of observed performance just as we do in assessing the acquisition of so-called "skills". The task we have set ourselves in the PACKS project, of trying to understand the reasons for children's actions in carrying out investigations in science classrooms, is, therefore, more similar to the task facing someone who wishes to explore children's "understandings" within a science content domain than might at first appear.

In any given science domain, whether the emphasis is on "knowing what" or "knowing how", we are interested in children's performance in relation to an idea of the "ideal" performance which we want to help the child eventually to achieve. This comes from the structure of accepted knowledge in the domain, or the method followed routinely by members of the relevant scientific community in carrying out the practical operation in question. In some cases, these target performances may be simplified versions of the accepted scientific knowledge or practice. But they are derived from, and depend on the existence of, ways of thinking, talking and acting which are consensually agreed within the practising community of scientists.

In other words, ideas about progression in science have a strong **normative** dimension. They are about the movement of the learner from an initial "novice" position towards (though not necessarily all the way to) a final position which is known in advance. However, knowledge about the desired end-point is not much help in curriculum planning; a more fully articulated model of progression in a domain will have to specify some of the intermediate stages between the novice position and the desired end point, and may also provide further guidance on the nature and causes of transitions from one stage to another.

It may be useful to classify models of progression in the following way:

Level 1: A model of progression at this level aims to specify a series of stages of understanding between a basic or "novice" level and a more "expert" one. These may be in a linear hierarchical sequence, or they may be organised as a more complex network. (The ten levels of the National Curriculum (DES 1991) are an example of such a model. A more fully articulated version of this model

is provided by NEAB (1992).) Traditionally such models were almost entirely *analytic* and *a priori*, based upon an "expert" analysis of the logical structure of the domain in question. The differences between successive stages were seen in terms of *rational principles*, such as greater explanatory content or bringing a wider range of phenomena under a single explanatory framework, or as necessarily building upon ideas at previous stages. The experience of the last twenty years of research into children's ideas in various science domains has shown us that models of domains based only on *a priori* reasoning may be seriously flawed. Empirical information from studies of the actual understandings of children about aspects of the domain may be necessary to test, refine and develop such models (see, for instance, CLIS 1992). The empirical approach cannot, however, replace the analytic one but can only supplement it; empirical studies depend on having some kind of analytic model to guide the enquiry. This becomes, essentially, a hypothesis which is tested against empirical data and rejected or modified in the light of the data. A developed level 1 model of progression in a domain is the product of an interaction between analytic and empirical approaches.

Such models may also have something to say, or will embody implicit messages, about the nature of the changes involved in moving between stages. A key distinction, noted by several writers, is whether the step involves knowledge "accretion" or "restructuring" (Rumelhart and Norman 1981). West and Pines (1985) reflect the same distinction in their terms "conceptual development" and "conceptual change", whilst Carey (1985) talks of "weak restructuring" and "radical restructuring".

Level 2: A model of progression at level 2 will seek to provide an account of the *reasons for changes* of understanding from one stage of a level 1 model to the next. An issue in constructing such a model is to decide what sorts of explanations for change will be allowed. At one extreme, a level 2 model may allow only *rational*, or *epistemological*, explanations for change. Changes might be accounted for as rational responses to new evidence which has been brought to the learner's attention. The well known conceptual change model of Posner, Strike, Hewson and Gertzog (1982) is essentially of this sort. In general, "cognitive conflict" models of conceptual change tend to emphasise rational grounds for changes of understanding. Piaget uses the idea of

"equilibration" as the fundamental *reason* for development on rational grounds; it is "the end of the line" for explaining why individuals seek to reduce conflict between existing schemata and new information. But level 2 models of progression might also use *psychological variables* such as: need-reduction, personality characteristics, social pressure, attention, motivation to achieve. Indeed a more recent revision of the Posner *et al.* model (Strike and Posner 1992) recognises the need to include a wider range of factors, including motives and goals, in describing a learner's "conceptual ecology" and the ways in which it may change.

At both levels 1 and 2, a further distinction can usefully be made between models which aim to describe the progression of *populations* and those which are concerned with the progression of *individual learners*.

Table 1 A classification of models of progression

	deals with **populations**	deals with **individuals**
a "map" of stages	level 1 population	level 1 individual
a "map" of stages plus an account of reasons for moves from stage to stage	level 2 population	level 2 individual

All curriculum planning is based (usually implicitly) on at least a level 1 population model. In its details, it may contain some level 2 elements, but still in population terms. Teachers in their own classrooms may, of course, use strategies which draw upon their knowledge of individual learners. The classifications above also have some implications for research on progression. "Snap-shot" studies of children's understanding in a domain can, at best, provide data for improving level 1 population models. Cross-age studies might be used to provide data for developing level 1 or level 2 population models. Short term longitudinal studies (such as pre- and post-instruction designs) may provide useful data on specific transitions within level 2 models and can, in principle at least, provide data relevant to both population and individual

models. Longitudinal studies of a cohort can also provide data on individual models at level 1 or 2.

In terms of this classification, the PACKS project sets out to provide a level 1 population model of progression in the understanding required to carry out a scientific investigation. The PACKS data may also bear, albeit speculatively, on some aspects of a level 2 population model. Before considering how such models might be constructed, it may be useful to explore in a little more detail the nature of scientific investigations in the classroom and of the understanding involved in responding to such tasks.

Scientific investigation

The focus of the PACKS project appears at first sight to be on a "performance" domain rather than an "understanding" domain (though it is important to bear in mind the point made earlier, that the similarities may be greater than often assumed). Carrying out an investigation is, however, clearly a complex performance, involving a series of distinct sub-elements which are linked together as a sequence of actions. This complexity distinguishes it from a more simple "skill" such as using a microscope. It is also clear that decisions about how to proceed draw on *understanding of the science domain* in which the investigation is set, and on *understanding of the procedures* which should be followed in a scientific investigation.

It may be important, at this point, to clarify what is meant here by "a scientific investigation". The current interest in investigations in science is largely a consequence of the position they hold in the National Curriculum in Science, where Attainment Target 1 (Sc1) is about the ability to carry out scientific investigations. Essentially an investigation is an open-ended task. The NCC INSET Pack for Sc1 (NCC 1991) begins by emphasising the difference between *investigations* and other sorts of practical work, in particular *illuminative practicals*, designed to illustrate an idea or theory, and to lead to a predictable conclusion which the learner can use to develop his or her understanding in a domain. By contrast, an investigation leaves the child to make decisions about what apparatus to set up, what to vary, what to measure,

and so on. And the result is not a fact or relationship which the child is required to remember; the learning is about the procedures followed in collecting data to answer the original question posed.

Neither the National Curriculum nor its supporting documentation, however, says explicitly what it takes a "scientific investigation" to be. Instead, some pointers are provided or implied by the statements of attainment (SoA) and various associated NCC publications. The SoA emphasise relationships between variables and use the number and type of variables involved as an criterion of increasing "task demand". The Non-Statutory Guidance (NCC 1989) strongly reinforces this "variables perspective" on investigations. Sc1 has been criticised by some for appearing to limit "scientific investigations" to explorations of the relationship between a dependent variable and one or more independent variables.

In the PACKS project, we have taken an "investigation" to mean a task (whether given or self-imposed) which involves the collection and use of empirical data (observations or measurements) to answer a question. This definition is very wide. It encompasses enquiries in the social sciences as well as the natural sciences, and might even be claimed to include historical enquiry. So a "scientific investigation" must meet some additional criteria; but it is quite difficult to say precisely what these are. The boundary between science and other forms of enquiry is a matter of perennial scholarly debate (see, for example, Bernstein 1983). Many of the issues involved need not concern us here, and to discuss them at length would be a diversion from the main theme of this paper. Two specific points, however, may be worth stating explicitly. First we would include technology/engineering tasks within the science domain because of the similarities of subject matter and of conventions about purposes and explanatory styles (causal, non-teleological, etc.). So "scientific investigations" in our usage may be of artifacts as well as of natural phenomena and materials. Second, "scientific investigations" include enquiries which are *not* about relationships between variables, although these remain an important sub-category. In fact the investigations we used in the PACKS project *are* of the variables type, as these do provide insights into the questions we are focussing on and are acceptable to teachers and pupils because of their obvious similarity to Sc1 type.

As the domain of interest of the project might be described as "empirical enquiry in the scientific mode", it may be important to say a little more about the relationship between empirical enquiry and scientific enquiry. Another strand of criticism of Sc1, particularly in its first (DES 1989) formulation, was that it took a rather naive empiricist view of science. Investigations cited as examples were set in atheoretical contexts, such as the waterproofness of cloths, or the absorbency of paper towels[1]. The Sc1 view appeared to give insufficient recognition to the fact that variables arise from a theoretical view of the phenomenon being investigated and that science is more than simply the collection and codification of empirical data. The revised version of Sc1 (DES 1991), by insisting that assessment of Sc1 be carried out using investigations which include three distinct strands (predict/hypothesise, observe/measure, interpret/evaluate), can be seen as an attempt to address this issue. This revision has, however, raised further problems at the practical level; there appears, for example, to be a rather small number of phenomena which can be the subject of genuine, theory-based predictions (as opposed to guesses or hunches) at school level *and* can be investigated empirically using school equipment.

It would require a much longer paper to explore these issues in detail and to propose a set of criteria to characterise "scientific investigations". For the moment, let us just say that scientific enquiry is empirical enquiry carried out within the ontological framework of the accepted scientific view and within a framework of methodological commitments, which may include tacit and conventional elements. Its purpose may be explanatory, or simply information collection. But because *all* scientific enquiry involves empirical enquiry (although the converse does not follow, and involves the difficult issues of demarcation referred to above), a certain level of competence in (and hence of understanding of) empirical investigation is a necessary, though not a sufficient, condition for carrying out a scientific investigation. Therefore, observing children's responses to empirical investigations may provide useful information

[1] These are atheoretical in the sense that the child is expected to treat the investigation as a purely empirical enquiry. Theoretical explanations of differences observed, or theoretical considerations which might lead to selection of objects or variables to investigate are well beyond the level of understanding of the child.

about aspects of their understanding which are necessary for carrying out a scientific investigation. And empirical investigations may provide a teaching medium for developing understandings which are necessary for carrying out scientific investigations.

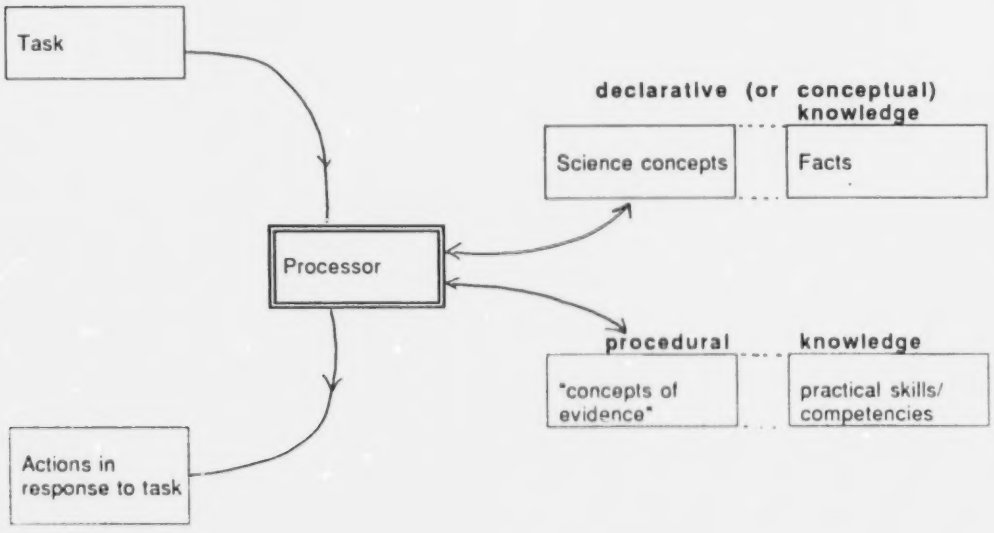
Performance of an investigation task in science

In order to explore children's performance in scientific investigations in relation to their understanding, we developed seven investigations, involving topics in biology, chemistry and physics. Small groups of children, at ages 9, 11 and 14, were closely observed as they carried out each of these investigations within a normal classroom setting. Each group was asked to write down their prediction about the outcome of the investigation, and their reasons for this prediction, before beginning. A researcher kept field notes on the performance of the groups observed and also conducted impromptu interviews with the children to clarify their reasons for the actions they were seen to be undertaking. These were tape-recorded. A more formal interview was conducted with each group after completion of the task. The children were then also asked to respond to probes (administered orally to 9 year olds and in written form to 11 and 14 year olds) designed to elicit their ideas about science content and investigation procedures relevant to the investigation they had just completed. The children's own written record of their investigation was also collected. This entire set of data was then used to produce a detailed case record of each group's performance on the investigation task, which included verbatim extracts from interviews.

In order to approach the question of progression, it is necessary to try to describe the stages of, and the understandings implicitly required for, a response to an investigation task. This is not a trivial task. In recent years, the introduction of practical assessment in GCSE Science, the advocacy of "process science", the work of the APU and the discussions surrounding National Curriculum Attainment Target Sc1 have produced a variety of terminology about the "knowing how" aspects of science learning. There is little agreement about terminology and no evidence of an agreed underlying model of the domain. This contrasts with the situation in most "knowing what" domains

in science, where working models *are* available, based on a mixture of *a priori* and empirical inputs (as discussed earlier), the precise composition of the mixture varying from domain to domain. A comparable model of the "knowing how" domain would be a major step forward and would provide a springboard for further work.

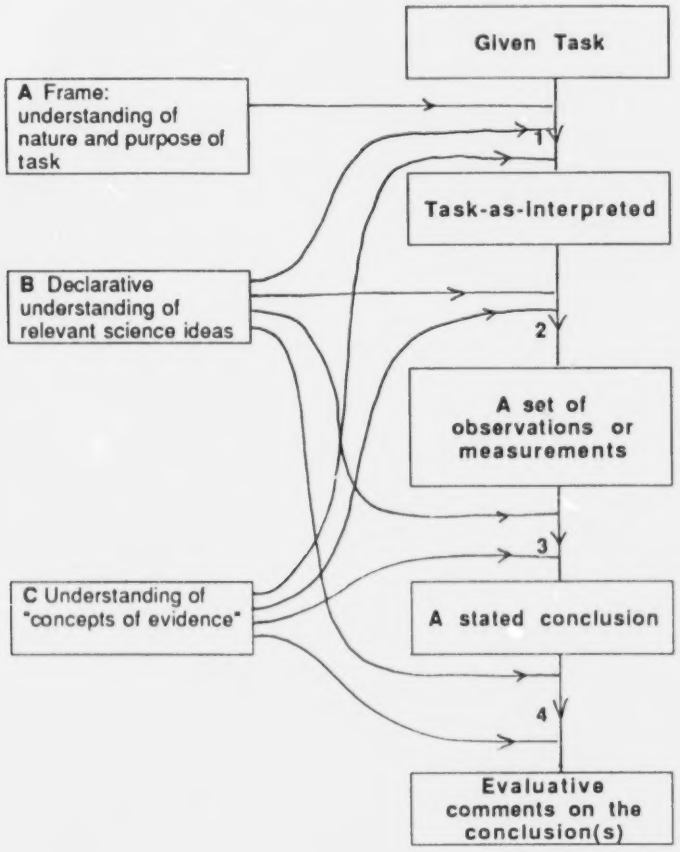
Figure 1 A simple model of the understanding which underpins investigation performance



A simple model of the thinking involved in investigating is shown in Figure 1. This indicates that, in order to "transform" the given task information into a sequence of actions (the "task performance"), some "processor" in the brain must select and access certain types of stored knowledge, and must integrate these into a plan of some sort which allows activity to begin. Figure 1 proposes (in line with the terminology of APU, 1987: 50ff) that the accessed knowledge might be usefully divided into two categories, the relevant *declarative (or conceptual) knowledge* about the science domain of the investigation, and *procedural knowledge* about investigating in science. These two categories have been further sub-divided to indicate a possible hierarchy within each.

This simple model does not, however, provide a basis for seeking to link specific aspects of understanding to specific stages in the performance of the investigation. For this, a model like that shown in Figure 2 is necessary.

Figure 2 Carrying out a scientific investigation task: stages of performance and "types" of understanding involved



Before providing a commentary on this model, it may be useful to make clear that it is the product of an interplay between analytic and *a priori* reasoning on the one hand and empirical data from children's actual responses to investigation tasks on the other. It has been refined as the project has progressed, in the light of the data collected. So that data cannot be used as a "severe test" of the model. A future project could, however, set out to test the model. It can, however, be used to validate the model by showing how it can accommodate and illuminate the varieties of performance and understanding observed.

The points which follow provide a commentary on the model of Figure 2:

1 The five boxes which summarise the stages contain *"objects"* - things which are produced (or may be produced) in the course of the investigation.

2 The numbered vertical arrows identify *actions*. The grounds for these are not directly observable. We can only infer them, either from the "objects" produced (in the sense above) or from the accounts children give to explain why they are doing what they are doing (or why they did what they did).

If we use the term "skill" we are referring to the quality of the actions represented by the numbered vertical arrows. This "quality" may be reflected in the nature of the decisions taken or of the specific actions performed in moving from one stage to the next.

3 The step shown by arrow 2 is sometimes referred to as *planning*, or *devising a strategy*. This must, however, be interpreted widely. Some children appear to from a complete plan for taking measurements/ observations before proceeding; others appear simply to proceed stepwise from one to the next. Intermediate levels of strategic planning are also possible: deciding on the first few measurements and doing this before deciding what to do next.

4 The stages of an investigation are shown in Figure 2 as a linear sequence. In practice, children may go back from a later stage to revisit an earlier one, perhaps in the light of trials and experience, or as a result of a re-think (see APU 1987: 24). Feedback loops of this sort have been omitted in Figure 2 for clarity only. The model should be read as including the possibility of such loops from any stage to any previous stage.

5 The distinctive feature of the model is that it identifies three different types of understandings which are brought to bear in taking decisions about actions. These we have called: *frame, declarative understanding* and *understanding of "concepts of evidence"*. The first and third of these together form *procedural understanding*.

- *Frame* refers to the child's understanding of the nature and purpose of the task. Some children interpret a *scientific* task (involving the relationships between variables) as an *engineering* task (optimising the effect of variables) (Schauble *et al.* 1991). We have also observed two less sophisticated frames, which we term *modelling* and *engagement*. In addition to this internal sense of frame, there is an external dimension which relates to the child's interpretation of the task as a classroom event. For the children we observed, the task we gave was, in effect, to carry out a *school science investigation* (rather than simply a *science investigation*). Children bring an understanding of a range of social norms and expectations to bear on their performance of such tasks. We cannot simply assume that they consider they are engaged in "finding out information"; they may, rather, be "satisfying the requirements of the social situation". (Strike and Posner (1992: 161) make a similar point in their discussion of science teaching for conceptual change.) Much of this understanding is tacit, but it is likely to have a significant influence on the performances observed.

- *Declarative understanding* refers to the child's understanding of facts, ideas and concepts in the science domain with which the investigation deals.

- *Understanding of "concepts of evidence"* refers to the child's understanding of criteria for assessing and evaluating the quality of empirical evidence. Three key notions are: *reliability* (can a datum be taken as a good measure of what it is measuring?); *structural validity* (does a set of data, if we assume each is individually valid and reliable, permit a logically valid conclusion to be drawn?); and *validity* (is a datum measuring what you think it is measuring?). Understanding the idea of a *variable*, and perceiving phenomena in terms of relations between variables, is also an important aspects of understanding within this category.

6 The model leaves open issues of the structure of the contents of the boxes labelled *declarative understanding* and *understanding of "concepts of*

evidence". It is likely to be useful to think in terms of different types of memory element (see, for example, White 1988: 22ff.) within both these categories. One important type is likely to be what White terms an *episode* and others (Schank and Abelson 1977) have called a *script*. Adoption of a "cooling curve" arrangement for our *Cool Drink* investigation (which is descibed more fully in the next section of this paper), and aspects of the use of the notion of a "fair test", for example, might be seen as scripts for the two types of understanding.

7 We are particularly interested in the ideas that children draw from each of these three sources to carry out each of the steps identified by the numbered arrows and how they use these ideas. The arrows linking the "understandings boxes" to the stages of action show where we believe ideas of each type might be brought to bear. It may appear, at first sight, that each type of understanding is involved at each stage. The way, however, in which each is involved differs at different stages.

So, for example, if we take a decision to repeat measurements to indicate an understanding of the notion of reliability, this might be evident at the planning stage (arrow 2), or the decision might be taken later in the light of events (as a loop back initiated during arrow 3). A decision about the choice of range of an independent variable, or about the value chosen for a control variable, is also an indication of an understanding of reliability, but is more likely to be built into the performance at the planning stage (arrow 2).

In order for a data set to have structural validity, this must either be built in at the planning stage, be added at a later stage as a "repair", or arise "by chance". It may not always be easy to tell which of these is the case as planning decisions are not always evident through explicit statements or actions. Structural validity also depends on the nature of the conclusions the child wishes to draw (which may be reflected in a reformulation of the given task into the task-as-interpreted). So, for example, a minimum of two measurements, with different values of one parameter and others held fixed, is necessary for drawing a conclusion later in terms of a *comparison*. It might be argued that a minimum of three is required to draw a conclusion in terms of a *trend*; and a minimum of three plus some discussion of

interpolation or extrapolation for a conclusion in terms of a *relationship*. The nature of these choices is of interest, as is the consistency of decisions at the planning stage (or, if no planning is apparent, of the data set which is in fact recorded) with the conclusions eventually drawn.

The complexity of investigation performance also means that, to a greater extent than with simpler types of task, *performance* is always likely to under-represent *competence*. A child may, for example, display an understanding, in a separate probe, of an aspect which was relevant to the investigation but not incorporated in his/her performance [1]. Such a child may be able to carry out an investigation, with understanding, if appropriately prompted by an "expert" at critical stages. Vygotsky's idea of the Zone of Proximal Development may be a useful resource in thinking about the implications of this situation.

Another way to approach this issue is to recognise that the *availability* of a memory element does not guarantee its actual *selection* for a given task. If a child provides an answer in a probe which appears to indicate understanding of a key idea, but has not made use of this idea in carrying out the investigation task, then this might imply that the idea is understood but that it has either not been selected, or not judged to be relevant, to the task. On the other hand, absence of use in the task might be taken as indicating a gap in understanding, and the answer to the probe seen as misleading. For Piaget, evidence of *development* is in the consistent ability to apply an idea when it is relevant to a problem. In a strict interpretation of this view, inconsistencies between task and probe performance would indicate lack of understanding of the idea in question. In the context of very open-ended tasks, however, it seems necessary to modify this interpretation by including some notion of "triggering" of ideas which the child understands. Failure to use an understood idea then raises questions

---

[1] This point, of course, applies to any "problem-solving" situation. Plato's classic account (Plato, [1956]) of Socrates leading the slave boy to the derivation of the square root of 2 shows how it is possible for someone to have all of the knowledge necessary to solve a problem, yet not to be able to select the right information and to combine it in the correct manner to reach a solution.

A converse point can also be made: that some performances may, on occasion, give a misleadingly high impression of competence or understanding. An answer based on rote learning might, for instance, be taken to imply understanding. To some extent, this is a question of the validity of the probes or tasks used to assess understanding or competence. It also embodies a view of what we mean by "understanding" or "competence", namely that it should be reflective and articulable in the possessor's "own words".

about the nature of the "trigger" in the task specification or context. Differences between the ideas children display in their performance of a task and in their answers to subsequent probes of understanding are clearly of interest in this regard.

An illustration of the model in action: choices of frame in the *Cool Drink* task

We will now use one aspect of the model of Figure 2 to illustrate how this model links to the more general ideas about progression discussed earlier in the paper. The aspect chosen is children's differing ideas about the purpose of an investigation, as evidenced by their responses to the investigation task *Cool Drink*. This task is set in the context of a cool bag used to keep a drink cool on a hot day. Children are asked to "find out how the thickness of the padding affects how well a cool bag works". For the task, groups of children are provided with a clock, a thermometer, and a range of containers such as beakers and measuring cylinders. They have access to a supply of ice-cold water, but the use of any type of tap water is not excluded. They also have access to a supply of three types of common wrapping materials: fleece, bubble wrap and foam. Each group is shown how they can vary the thickness of the wrapping by using different numbers of layers.

In children's responses to the task, we noted examples of four distinct frames, or conceptualisations of the task, three of which had sub-categories:

Engagement frame	1	Engagement and activity with the apparatus provided without obvious plan or purpose
Modelling frame	2A	Modelling to produce a desired physical appearance
	2B	Modelling to produce a desired effect
	2C	Modelling to produce a desired phenomenon
Engineering frame	3A	Random engineering: optimising the desired effect by trial and error
	3B	Iterative engineering: seeking a combination of factors which optimise the desired effect
Scientific frame	4A	Scientific comparison: using 2 values (often extreme values) of independent variable
	4B	Scientific range: using 3+ values of independent variable to identify a trend
	4C	Full scientific: using 3+ values of independent variable to find a functional relationship

Frames 1 to 3A might be classed as novice frames. The higher levels of frame require planning and use of data for further investigation and so contain some more "expert" aspects.

The *Cool Drink* task is open to interpretation in terms of a variety of frames. The given task asks children to explore the relationship between two variables: the thickness of the padding and the temperature change. The independent variable (thickness) is a continuous variable, which is available in discrete values. So the task is posed within a scientific frame. Many groups, however, re-formulate the task in an alternative frame, or use several different frames during the task.

Instances of an engagement frame

Use of the engagement frame is characterised by an immediate and sustained engagement with the equipment provided, but without apparent plan or purpose. Investigating appears to be characterised by activity - the pressing need of the children seems to be to "do something", allied to an apparent belief that this action will satisfy the demands of the task. In the *Cool Drink* task, the engagement frame was relatively rare. It did, however, appear in the determination of some groups to use of all the different types of containers provided. This was found more frequently with Y4 children than with Y7 or Y9. One Y4 group insisted on using different types of water which were not specifically offered but were available (hot and cold tap water, and ice-cold water), but went on to use the type of water as the independent variable in an investigation with an engineering frame, to find which water keeps coolest in the beaker.

Instances of a modelling frame

Several pieces of evidence were taken to indicate a modelling frame focussing on the physical appearance of the cool bag (frame 2A). One is an explicit justification of the choice of the measuring cylinder as the container in terms of its similar shape, e.g. "because it looks more like a cool bag" (stated by several Y7 and Y9 groups). Similarly, the comment that "we should have tested it on a proper drink bottle or flask" (a Y9 group) and the suggested improvement of "construct[ing] a proper cool bag" (Y9 group) indicate a modelling frame 2A.

Another indicator of frame 2A is the indiscriminate wrapping of the container in padding material. Typical pupil comments associated with this response are: "We have made a proper cool bag. It [the bubble wrap] is all over." "Cool bags are always like that. They are all covered. Everywhere. To keep the food fresh, and the drink fresh." (Y9 group).

A broader version of the 2A frame can be seen in one Y9 group's reason for their choice of observation period: "We tried it [measuring the temperature change] for 2 minutes and it stayed the same. So we'll try it for longer." "Yes, because if you have a cool drink, you are not going to put it in, and then drink it after 2 minutes, are you?" In a similar vein, a Y7 group suggested an improvement of their investigation would be "doing the experiment outside as the school is warm anyway, so it ruins the results".

Frame 2B differs slightly from this in that the aim is to produce a desired effect, in this case a decreasing or low temperature. One indicator of frame 2B is when groups plan to work with only one set-up (so with no intention to compare), and want only to establish whether their set-up "works" (examples from Y4, Y7 and Y9). One Y7 group explained that they "want to keep testing the temperature with the thermometer. We hope it will stay the same temperature". Another planned "to put plenty of popping paper round the outside to keep it [the temperature] how we want it". A Y9 group wrapped a beaker of ice-cold water in a thick layer of foam "just [to] see what happens to the temperature".

Some groups show that they are working within a 2B modelling frame by treating readings which do not show a constant or decreasing temperature as indicating "failure" or proof that "the experiment doesn't work" (groups at Y4, Y7 and Y9), even to the extent that no conclusions are drawn and only readings are reported. Conversely, but within the same frame, some other groups use a subjective, unspecified benchmark for concluding that the chosen set-up was "very good, because the temperature didn't change very much" (Y7 group), did confirm their prediction because "the temperature did not go up very much [8 degrees in 30 minutes!]" (Y9 group) or "went up very slowly " (Y9 group), or that "the thick padding kept the [cold] drink just right" (Y9 group).

If groups working within frame 2B feel they have failed at their first attempt, most of them try a second set-up. Few, however, do this as a comparison, which would move

them into frame 3A. The majority see their new attempt as an independent, new test to see if 'this will work'. Several methods are used to encourage the temperature to go down: add more water, or re-arrange the 'cold' material (a Y4 and a Y9 group), test another 'cold' material (Y7 group), stuff the 'cold' material inside the beaker (Y4 group), soak the padding material around the container in ice-cold water (Y4 group), use new or additional (cold) water (Y4 and Y7 groups), use different types of water, including hot tap water which provides a satisfactory drop in temperature (Y4 group).

Instances of an engineering frame

A random engineering frame (3A) is used frequently by groups of all three ages. It occurs when groups decide to "find out which type of material is the best one to keep the liquid cool". The same frame is being used when combinations of materials are compared for effectiveness (instances at Y4, Y7 and Y9). In both cases, this is a complete re-formulation of the given task, focusing on type of material rather than thickness of material as the independent variable.

A similar engineering frame may be used with thickness as the independent variable: the task is conceptualised as finding the thickness which makes the temperature fall most (groups at Y4 and Y9) or fastest. One Y9 group used 4 beakers with increasing thickness of foam, intending to "time how long it takes for each beaker to drop in temperature and then see which works best".

Some groups show that they are using a random engineering frame (3A) when asked to draw conclusions. These are formulated in terms of an optimum. In fact, if the type of material is chosen as the independent variable this forces the investigation into a random engineering frame. Groups may look for the material which keeps the temperature lowest or constant (examples at Y7 and Y9). Some even seek a material which makes the temperature decrease: "foam and fleece are the best because they go down." (Y7 group); or decrease quickly: "felt is the best because it is the fastest." (Y4 group). Other groups use the time as the independent variable and select the material which takes longest to allow a rise in temperature (groups at Y7 and Y9).

An iterative engineering frame (3B) occurs rarely in the *Cool Drink* investigation and might be expected to arise more commonly in an investigation with two or more

independent variables. A few groups, however, use the type of material as a second variable in the manner illustrated by this extract from the case report on a Y4 group:

> In the first trial they measured the temperature for each beaker wrapped in one layer of the different materials: fleece, foam and bubble wrap. After 10 seconds they read the thermometer, and it is clear to them that the fleece kept the water at five and a half degrees, and bubble wrap and foam at six degrees, so their favourite was fleece. The girls reasoned that "the one layer of fleece gives most cold to the beaker, so more layers will make it even colder". As fleece seemed the "best" material, the girls' second trial measured the temperature (after 10 seconds again) for a beaker wrapped in two and three layers of fleece.

Another Y4 group used an iterative engineering frame by taking the type of water as a second variable. They initially compared the temperatures of different mixtures of tap water finding 16, 24 and 32 degrees. Having found the water with the lowest temperature, they then compared the temperature change of this water in containers wrapped in the various materials.

Instances of a scientific frame

A scientific comparison frame (4A) is characterised by the use of two thicknesses and a conclusion drawn in the form of a comparison (or even a trend) (groups at Y4, Y7 and Y9). This example is taken from a Y9 group:

> For their first trial the boys measure consecutively the temperature at 1 minute intervals over 5 minutes for 2 beakers with 100 ml of cold tap water, one wrapped in one the other in two layers of fleece "to see if two layers keeps it cooler." The starting temperature is reported as 14 degrees for both beakers. The results (in table form) show an increase over 5 minutes from 14 to 16 degrees with one layer, and from 14 to 15 degrees with two layers. The boys conclude that "the thicker the insulation the longer the water stays cool". They generalise from a 2-point comparison.

Alternatively, but still a scientific comparison frame (4A) two different materials are used for tests but the results are seen as due to the thickness of the materials, allowing again for a conclusion in the form of a trend.

> The group tests the temperature change for a period of 3 minutes for a beaker wrapped in foam (= thin) and an other one in fleece (= thick). In the first case they find the temperature to remain at 4 degrees; in the second case the temperature has increased from 5 to 6 degrees. The group concludes that "the thinner it is, the cooler it keeps it". (a Y4 group)

A scientific range frame (4B) is used when groups use more then 2 data points and draw conclusions in the form of a trend (many examples at Y7 and Y9). This example is the approach used by a Y9 group:

> As the first trial the boys plan to use 10 ml of ice-cold water in a cylinder wrapped in one layer of foam, and measure the temperature every minute for 10 minutes. They include a reading at the beginning. They use the same procedure for a measuring cylinder without any padding, and with 2 and 3 layers of foam respectively.
> The final temperatures decrease with increasing numbers of layers of padding.
> They conclude that "the more padding you put on, the cooler it stays" or "the more insulation, the longer it takes to warm up".

Stability of frame

As has been indicated in the discussion above, some groups appear to alter frame during an investigation, or to use a different frame in carrying out the task and in stating conclusions. Three-quarters of all groups, however, used a stable frame (frame 2, 3 or 4) during the investigation. Frame as a description of children's perception of 'what an investigation requires' seems relatively fixed and is only marginally affected by what happens during the investigation.

It is also interesting that over 75% of the changes of frame which occur are towards an engineering frame. This is an improvement if a group starts from a modelling frame, and we could claim that learning has taken place during the investigation. Several groups, however, also regressed from a scientific frame to an engineering frame.

Frame choices at different ages

In the PACKS research design, age is seen as only a very crude predictor of performance level. Even so it is of interest to look at choices of frame for the *Cool Drink* investigation by children of the three different ages involved (Table 2). Where a group used more than one frame during the task, the frame they used when formulating their conclusions has been taken to indicate their 'main' frame.

Table 2 Choice of investigation frame with age

'main' frame	Y4	Y7	Y9
modelling frame	38%	20%	20%
engineering frame	38	55	40
scientific frame	24	25	40
	100	100	100

This table shows a clear shift between Y4 and Y7 away from the modelling frame and towards the engineering frame. Another change occurs between Y7 and Y9, from the engineering frame to the scientific frame. Even at Y9 level, however, fewer than half of the groups used a scientific frame for a scientific task, and the choice of an engineering frame was still very common. It would be particularly interesting to see if older pupils showed a more clear cut shift towards the scientific frame.

Implications of choice of frame

As has already been said, the *Cool Drink* task as given asks for a response within the scientific frame. The aim of the task is to find out how the thickness of insulation affects the rate of temperature rise of a cold liquid. Many groups of pupils appear to reformulate the task in a different frame at a very early stage in their investigation, in making the step identified by arrow 1 in Figure 2. In some cases, this may result simply from a mistaken reading of the task, due either to haste or carelessness. The provision of several materials (seen by us as a means of increasing the options available) may have acted for some as a distraction, encouraging the choice of type of material as the independent variable (IV). Once a group has opted for type of material as IV, then their investigation is almost inevitably set within what we have termed an engineering frame.

There is, however, another possible interpretation of this reformulation of the task in other than a scientific frame. It is that children interpret a given task in a form in which they believe they can succeed. In other words, if a given task appears too demanding, it is reformulated, almost intuitively, as one which is

feasible and manageable for that individual or group; the group is not aware that it has altered the task from that given. The task-as-interpreted is then one which is understandable to the group and their resulting performance is consistent with this self-imposed task. If so, then the step of modifying the given task into the task-as-interpreted may be a very significant one in making sense of subsequent actions and decisions.

In any event, the large number of children who choose frames other than the scientific one for this investigation suggests that more explicit teaching about the purpose of scientific investigations may be required, emphasising the difference between modelling and engineering (or optimising) approaches and the scientific intention of understanding a phenomenon, often in terms of relationships between identified variables.

What can we say about progression?

The preceding discussion of choices of frame is intended as an illustration of the possible nature of progression in one of the aspects of understanding identified in the model of Figure 2. In this final section, we will review some more general issues about progression in the understanding which underpins performance of scientific investigations.

The central problem in assessing students' work on science investigation tasks and in specifying the nature of progression in performance stems from the lack of clarity about the nature of the understandings involved in producing the wide variety of performances observed. As a result, assessments of performance have tended to be either:

i Global evaluations of the overall quality of students' work in response to a task. The complexity of the performance required in response to an investigation task, however, means that global measures of performance (scores) risk being unduly impressionistic. They are, in any case, not particularly helpful since the same global "score" can be arrived at in very different ways; the relative weighting of different elements is necessarily

subjective; "failures" at one stage may foreclose the possibility of certain decisions or options at later stages.

ii Fragmented and piecemeal evaluations of discrete elements of investigation "skill". These leave the sense that the whole is appreciably greater than the sum of the parts and that the essence of investigation performance is not "captured" by assessments of separate elements.

iii Focused rather narrowly on one aspect of investigation performance. The National Curriculum's focus on variables handling might be seen as an example of this, leading to a relative neglect of other important aspects and a consequent effective narrowing of the definition of an "investigation" in practice.

The value of the model of Figure 2 is that it proposes a small number of distinct aspects of understanding which underpin performance. Our data, though there is not space here to provide evidence to support the point, suggest that understanding in Box 2 (*declarative understanding of relevant science concepts*) can be treated independently of understanding in Boxes A and C (*frame* and *understanding of "concepts of evidence"*). We have, for example, cases where groups appear, in their investigation performance or in the probes of understanding, or both, to have a good understanding of the science which relates to the task; but use a "novice" frame or show a poor understanding of one or more of the key "concepts of evidence". Conversely, we have cases where groups with a distinctly "alternative" model of the science involved in the task nonetheless collect data and/or respond to probes in a way which appears to show good understanding of "concepts of evidence" within a more "expert" frame.

If this is so, then Boxes A and C of Figure 2 may provide a means of characterising procedural understanding in science at school level. We might use these categories to assess the performance of individual learners and to produce a profile of their understanding. The elements of such a profile would include:

i understanding of the purpose of scientific investigations (as indicated by decisions about *frame*);
ii competence in specific manipulative skills, including appropriate and safe use of instruments and equipment;
iii understanding of key "concepts of evidence", such as the idea of *a variable*, and the concepts of *reliability, validity* and *structural validity*.

Finally, and rather more speculatively, we might then raise some questions about the nature of the learning involved in moving from "novice" to more "expert" understanding and hence performance. Whilst it is clear that learners' decisions about frame may depend on the actual investigation task they are given, there appears to be a hierarchy of frames. Whilst some of these ideas about purpose of scientific investigation may be picked up tacitly, there may also be a case, as we have said above, for a more explicit discussion of these issues at regular intervals during a child's school science programme.

The "concepts of evidence" category deals with the understanding of ideas which relate to almost all empirical enquiry. It is reasonable to suggest that this is a domain about which all children, however "novice", know a great deal. Every person must, from infancy, have well developed skills for finding out information about the natural world, simply in order to make sense of experience. So we should expect children to have some well developed "prior ideas" (prior, that is, to formal instruction) about many aspects of our *understanding of "concepts of evidence"* box. For instance, the realisation that repeat measurements or observations may yield different results, especially with biological systems, could quite plausibly arise from everyday experience, though this is unlikely to be formulated explicitly as a general principle. It may even be that many of the fundamental aspects of understanding of "concepts of evidence" are available for use from an early stage, perhaps even from infancy, but that the ability to reflect upon them explicitly and the development of the language to talk about them comes later.

On the other hand, some ideas, such as the idea of a variable, and of seeking a relationship between a pair of variables (as opposed to a comparison between two cases or a trend in the influence of a factor on an outcome), are rather specialised ways of thinking, which are specific to the scientific view, and

unlikely to arise simply through experience. An understanding of this aspect of what it means to do a scientific investigation is necessarily learned. Other ideas about variables, in particular an understanding of the need to control variables and alter one at a time in order to be able to associate changes with a specific factor (the *structural validity* element of our *understanding of "concepts of evidence"* box) are regarded, within the Piagetian framework, as developmental and acquired as part of the achievement of formal reasoning. This interpretation is not, of course, universally accepted and younger children appear, in some cases, to be able to understand the need to hold all other factors constant in order to explore the effect of one factor which is varied.

We can, therefore, distinguish at least three different types of account which might be used to describe and account for progression in understanding and performance of science investigation tasks:

- a **domain specific knowledge-based** account. The domain is the understanding of "concepts of evidence".

- a **language competence** account. This would take many (perhaps all) the necessary competences to be available to the child, but would see the application of these to specific cases as limited by the extent of their ability to talk about, and hence to reflect upon, their understanding in order to see how to apply it to the case in hand.

- a **developmental** account. This would apply particularly to the aspect of performance associated with variables and structural validity.

The last of these has been much more widely discussed in the research literature than the others. We would suggest that, whilst all three may play some part in accounting for aspects of progression in investigation performance, the first two types of account merit greater attention. In particular, the overwhelming impression of many of the PACKS case records is that lack of understanding of what we have termed "concepts of evidence" is the major obstacle to satisfactory investigation performance for most children.

Acknowledgements

This work is supported by the Economic and Social Research Council grant no. X208 25 2008.

References:

Assessment of Performance Unit (APU). (1987). *Assessing Investigations at ages 13 and 15. Science Report for Teachers: 9.* London: DES.

Bernstein, R.J. (1983). *Beyond Objectivism and Relativism.* Oxford: Blackwell.

Carey, S. (1985). *Conceptual Change in Childhood.* Cambridge, MA: MIT Press.

Children's Learning in Science Research Group (CLIS) (1992). *Leeds National Curriculum Science Support Project.* Leeds City Council, Department of Education and CLIS Research Group, University of Leeds.

Department of Education and Science (DES) (1989). *Science in the National Curriculum.* London: HMSO.

Department of Education and Science (DES) (1991). *Science in the National Curriculum (1991).* London, HMSO.

National Curriculum Council (NCC) (1989). *Science. Non-Statutory Guidance.* York: NCC.

National Curriculum Council (NCC) (1991). *NCC INSET Resources. Science Explorations.* York: NCC.

Northern Examinations and Assessment Board (NEAB) (1992). Science Framework. Common Themes. Syllabus for the GCSE 1994 Examination. Manchester: NEAB

Plato ([1956]). *Meno,* 81E-86C. trans. W.K.C. Guthrie. Harmondsworth: Penguin.

Posner, G., Strike, K., Hewson, P. and Gertzog, W. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education,* 66 (2), 211-227.

Qualter, A., Strang, J., Swatton, P. and Taylor, P. (1990). *Exploration. A Way of Learning Science.* Oxford: Blackwell.

Rumelhart, D.E. and Norman, D.A. (1981). Analogical processes in learning. In J.R. Anderson (ed.), *Cognitive Skills and their Acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Ryle, G. (1949). *The Concept of Mind.* Harmondsworth: Penguin.

Schank, R. and Abelson, R. (1977). *Scripts, Plans, Goals and Understanding. An Enquiry into Human Knowledge Structures.* Hillsdale, NJ: Lawrence Erlbaum.

Schauble, L., Klopfer, L.E. and Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching,* 28 (9), 859-882.

Strike, K.A. and Posner, G.J. (1992). A revisionist theory of conceptual change. In R. Duschl and R. Hamilton (eds.), *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice,* pp. 147-176. Albany, NY: SUNY Press.

West, L. and Pines, A. (eds.) (1985). *Cognitive Structure and Conceptual Change.* London: Academic Press.

White, R. (1988). *Learning Science.* Oxford: Blackwell.